

REMARKS

Applicant has filed herewith a Request for Continued Examination. This Amendment address the rejections in the Final Office Action and adds new claims 25-37.

In the Final Office Action, claims 17-24 were rejected as being unpatentable under 35 U.S.C. §103(a) Benyassine in view of U.S. Patent No. 6,381,570 to Li et al.

The rejections are respectfully traversed in view of the following discussion.

I. THE CLAIMED INVENTION

The present invention, as disclosed in the claims and specification, relates to a method for improving the estimation of background noise versus voice energy in a voice activity detection (VAD) circuit. A method of the preferred embodiment for improving the prior art VAD circuit is implemented to improve circuits already using the ITU G.729 Annex B standards for *Silence Compression Scheme* (see enclosed copy of ITU G.729 Annex B submitted for the Examiner's convenience). The recommendations in the G.729 standard contains three different algorithms: 1) Voice Activity Detection (VAD), 2) Discontinuous Transmission (DTX), and 3) Comfort Noise Generator (CNG) generator algorithms, each serving different purposes of the silence compression scheme.

In the initial decision process, the VAD algorithm extracts parametric information from a received frame and makes a difference measure of the parametrics from the frame with running averages of the same parametrics in background noise encountered in the transmission. As a result, if the VAD algorithm detects voice activity, then the frame is sent on to a speech decoder 3 (see Figure 1 of the Application). If the VAD algorithm does not detect voice activity in the frame, then the frame is declared as "non-active voice" and sent to DTX/CNG algorithms used to code/decode the non-active voice frame (See App. pp. 3-4). The DTX algorithm decides if a set of non-active voice update parameters ought to be sent to the speech encoder, by measuring the changes in the non-active voice signal.

After an initialization, the running averages for the background noise characteristics are updated only in the presence of background noise and not in the presence of speech frames (App., p. 4, line 21). As shown in Figure 4 of the Application and in Figure B.2/G.729 "VAD flowchart" in the G.729 Annex B standard page 3, this update of *the running averages for the background noise characteristics occurs in the VAD algorithm*, not in the DTX algorithm. The standard states "The running averages of the background noise characteristics are updated at the last stage of the VAD module." (G.729 Annex B, p. 8) This is also explained in Benyassine, page 67 in the eighth paragraph beginning with "Updating the Running Averages of the Background Noise Characteristics."

The problem solved by the present invention is a problem within the G.729 Annex B VAD algorithm, which is that circumstances commonly cause the running averages to substantially diverge from the background noise characteristics of current and future frames (see Application, p.5). When the correlation diverges, the VAD algorithm has increasing difficulty distinguishing between frames of noise and frames of voice. The VAD algorithm reaches a point at which it can no longer tell the difference between noise and voice, stops updating the average running background noise, and interprets all the remaining incoming frames as voice frames, thus eliminating the bandwidth savings intended by the use of the G.729 Annex B standard.

The present invention intervenes in the VAD algorithm before it reaches a point where it stops discriminating between noise and voice and interprets all frames as voice. The present invention claims an additional, "*supplemental*" running average of the background noise parameters that is generated separately from, but at the same time as, the running average used by the G.729 Annex B VAD algorithm. When the claimed invention detects that the G.729 Annex B method for updating the running average has diverged from the that the supplemental running average of background parameters, then the supplemental running average is substituted into the VAD algorithm in place of the standard G.729 Annex B running average.

Thus, "by substituting the supplemental algorithm's characterization of the background noise for that of the G.729 Annex B algorithm, the estimations of noise and voice energy may be decoupled and made independent of the G.729 Annex B characterization when divergence occurs." (App., p. 9, line 20-24).

II. THE BENYASSINE AND LI REFERENCES

The Examiner has rejected claims 17-24 as unpatentable over Benyassine in view of Li. Applicant submits, however, that there the elements of the claimed invention are neither taught nor suggested by Benyassine in view of Li.

Applicant respectfully submits that the Examiner's rejections are incorrect and should be withdrawn for at least three reasons. First, the Benyassine passages referenced by the Examiner are merely summaries of the ITU G.729 Annex B standards without adding anything new and does not teach or disclose the claimed invention. Second, the text of claims recited by the Examiner in the Final Office Action misquote the text of Applicant's claims and do not sufficiently summarize or reflect Applicant's claim language such that the Examiner's language could be used to reject Applicant's claims. Third, the Examiner has primarily used page 68 of Benyassine to reject Applicant's independent claim 17. Page 68 describes the DTX algorithm and its components. This is a different part of the G.729 Annex B scheme than the VAD algorithm improved by the present invention, and the DTX algorithm description provides no framework for rejecting the present VAD algorithm improvement.

The Benyassine reference is nothing more than a summary of the ITU G.729 Annex B standards in the first six pages in combined with a report of MOS test results from a listening test using the G.729A codec in the last four pages. In contrast, the present invention is a new, supplemental method to improve the G.729 Annex B VAD algorithm. In the Abstract of Benyassine, the text begins by describing what is found in the text:

This article describes the recently adopted Annex B to the ITU-T Recommendation G.729. Annex B defines a low-bit-rate silence compression scheme designed and optimized to work in conjunction with both the full version of G729 and its low-complexity Annex A.

Thus, Benyassine introduces no new algorithms or supplemental methods for implementing or improving the G.729 Annex B VAD algorithm beyond what existed in the standards and does not teach or disclose "substituting supplemental average background noise parameters derived according to a supplemental algorithm for a running average of background noise parameters derived according to G.729 Annex B" as recited in claim 1.

The text of claims recited by the Examiner in the Final Office Action misquote the text of Applicant's claims and do not sufficiently summarize or reflect Applicant's claim 17 language such that the Examiner's language could be used to reject Applicant's claims. It appears that the Examiner has used the text of canceled claim 5 as the language in the basis of rejections for claim 17. In paragraph 2 of the Office Action, the preamble recited by the Examiner in rejecting claims 17 and 18 states:

...the method of converging an ITU Recommendation G.729 Annex B voice activity detection (VAD) device, comprising the steps of:

The Examiner's preamble leaves out nearly all of Applicant's preamble language. The language missing from Examiner's preamble is italicized below in the copy of claim 17's preamble:

A method for improving estimates of average background noise energy in a G.729 Annex B compliant voice activity detection (VAD) device by substituting supplemental average background noise parameters derived according to a supplemental algorithm for a running average of background noise parameters derived according to G.729 Annex B, comprising

In fact, the Examiner does not even mention background noise energy as used in the

running average of background noise parameters in the VAD algorithm anywhere in the rejection. Further, Benyassine is a summary of "a silence compression scheme," which the Examiner does not mention. Thus, the rejection does not address the claimed invention nor provides a framework for rejecting Claim 17.

Further, the Examiner has recited generic terms of "comparing a number of energy measures of a signal to said noise threshold value ... wherein only the energy measures of said number of energy measures having values less than said noise threshold value..." and "... determining a second value representing an average of said number of energy measures.... substituting said first value for said second value when a specific event occurs." The use of "energy measures" is too vague and indefinite a term to be used in a rejection of a claim for "supplemental average background noise parameters" that are substituted for the "running average of background noise parameters" in the *VAD algorithm*, as recited in claim 17. Applicant does use the language "a number of energy measures" anywhere in the claim. Using terms from Benyassine, there are numerous "energy" from signals in the three different algorithms described for G.729 Annex B:

In rejecting claim 17, the Examiner is comparing aspects of the DTX Algorithm on pages 67-68 of the reference, which does not teach or suggest Applicant's claims of an improvement to the *VAD algorithm* of the G.729 Annex B. The VAD algorithm and the DTX algorithm are two separate modules of the G.729 Annex B that operate for different purposes and cannot be used in comparison of one to another. The DTX algorithm is for:

...a discontinuous transmission module measures the changes over time of the inactive voice signal characteristics and decides whether a new silence information descriptor frame should be sent to maintain the reproduction quality of the background noise at the receiving end. (see Benyassine, Abstract)

The Examiner alleged that “determining a first value representing an average of said number of energy measures, when said energy measure is less than said noise threshold, wherein only the energy measures of said number of energy measures having values less than said noise threshold value are used to determine said first value, determining a second value representing an average of said number of energy measures, and substituting said first value for said second value when a specific event occurs teaches the claimed invention. However all these rejections are drawn from Benyassine’s summary of the “DTX Algorithm,” which is a separate and different module of the G.729 Annex B recommendation than the VAD algorithm that the present invention improves. For example, the Examiner cites “energy measures” as disclosing “running average of background noise parameters” in claim 17. However, the “energy measures cited on paragraph 2 of page 68 in Benyassine are for “residual energy” in the in the DTX algorithm, which is not comparable to the VAD algorithm’s background noise characterizations.

Finally, Applicant’s supplemental VAD algorithm and the G.729 Annex B VAD algorithm are “preferably separate entities that [are] executed in parallel...” (Application, p. 14, line 14). Benyassine only discusses the G.729 Annex B VAD algorithm and cannot possibly teach or disclose the claimed supplemental algorithm.

THE LI REFERENCE

The Examiner admitted that Benyassine does not disclose the method in the step of “generating a noise threshold” step of claim 17 and that Li discloses this step. However, Li fails to make up for the deficiencies of Benyassine, and the alleged combination does not teach or suggest the claimed invention. Li already mentions G.729 Annex B in its background (Col. 1, line 30), and Li is an improvement in VAD techniques of voice/noise discrimination. Li states “the system distinguishes active signal (e.g., voice, speech, etc.) from background noise to allow for the compression of or elimination of periods of silence or background noise. (Col. 2, lines 35–40). In other

words, if Li is applied to G.729 Annex B then it is an improvement to the VAD algorithm tasks of "voice activity decision." (Benyassine, p. 64 paragraph 1). Li would not be combined with the DTX algorithm methods of G.729 Annex B, as alleged by the Examiner and thus there is no basis for combining the two references. Further, the combination does not teach or suggest the claimed invention. Li, columns 7-8, describes a signal/noise discriminator. The claimed invention is for a supplemental average running background parametric. The two disclosures are separate parts of the VAD algorithm. Li's disclosure takes place earlier in the VAD decision process than the claimed invention's method of "substituting the supplemental average background noise parameters of the current period for the running average of the background noise parameters derived according to G.729 Annex B," as recited in claim 17.

For the foregoing reasons, the combination of Benyassine and Li do not teach or suggest the present invention as claimed in claims 17-24. Applicant respectfully requests the rejections be reconsidered and withdrawn.

III. CONCLUSION


In view of the foregoing, Applicant respectfully submits that claims 17-37, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above-identified Application to issue at the earliest possible time.

Should the Examiner find the above-identified Application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes that may be deemed advisable in a telephonic or personal interview. The Commissioner is hereby authorized to charge any fees associated with this communication to Client's Deposit Account No. 20-0668.


Appl. No. 09/871,779
Amdt dated August 17, 2005
Reply to Final Office Action of June 10, 2005

Respectfully submitted,

Date: 8-17-05


Kendal M. Sheets, Reg. No. 47,077
Joseph J. Zito, Reg. No. 32,076
Customer No. 23494
Local Telephone: (301) 601-5010

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: the Commissioner for Patents, United States Patent and Trademark Office, PO Box 1450, Alexandria, Virginia 22313-1450 on August 17, 2005.


Kendal M. Sheets
8-17-05
Date



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.729

Annex B

(11/96)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA

Digital transmission systems – Terminal equipments –
Coding of analogue signals by methods other than PCM

Coding of speech at 8 kbit/s using conjugate
structure algebraic-code-excited linear-prediction
(CS-ACELP)

**Annex B: A silence compression scheme for
G.729 optimized for terminals conforming to
Recommendation V.70**

ITU-T Recommendation G.729 – Annex B

(Previously CCITT Recommendation)

ITU-T G-SERIES RECOMMENDATIONS
TRANSMISSION SYSTEMS AND MEDIA

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
INTERNATIONAL ANALOGUE CARRIER SYSTEM	
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA CHARACTERISTICS	G.600–G.699
DIGITAL TRANSMISSION SYSTEMS	
TERMINAL EQUIPMENTS	G.700–G.799
General	G.700–G.709
Coding of analogue signals by pulse code modulation	G.710–G.719
Coding of analogue signals by methods other than PCM	G.720–G.729
Principal characteristics of primary multiplex equipment	G.730–G.739
Principal characteristics of second order multiplex equipment	G.740–G.749
Principal characteristics of higher order multiplex equipment	G.750–G.759
Principal characteristics of transcoder and digital multiplication equipment	G.760–G.769
Operations, administration and maintenance features of transmission equipment	G.770–G.779
Principal characteristics of multiplexing equipment for the synchronous digital hierarchy	G.780–G.789
Other terminal equipment	G.790–G.799
DIGITAL NETWORKS	G.800–G.899
General aspects	G.800–G.809
Design objectives for digital networks	G.810–G.819
Quality and availability targets	G.820–G.829
Network capabilities and functions	G.830–G.839
SDH network characteristics	G.840–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
General	G.900–G.909
Parameters for optical fibre cable systems	G.910–G.919
Digital sections at hierarchical bit rates based on a bit rate of 2048 kbit/s	G.920–G.929
Digital line transmission systems on cable at non-hierarchical bit rates	G.930–G.939
Digital line systems provided by FDM transmission bearers	G.940–G.949
Digital line systems	G.950–G.959
Digital section and digital transmission systems for customer access to ISDN	G.960–G.969
Optical fibre submarine cable systems	G.970–G.979
Optical line systems for local and access networks	G.980–G.999

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION G.729 – Annex B

A SILENCE COMPRESSION SCHEME FOR G.729 OPTIMIZED FOR TERMINALS CONFORMING TO RECOMMENDATION V.70

Summary

Annex B to G.729 defines a voice activity detector and comfort noise generator for use with G.729 or Annex A optimized for V.70 DSVD applications.

Source

Annex B to ITU-T Recommendation G.729, was prepared by ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 8th of November 1996.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had/had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 1997

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

	Page
B.1 Introduction	1
B.2 General description of the VAD/DTX/CNG algorithms	1
B.3 Detailed description of the VAD algorithm	2
B.3.1 Parameter extraction.....	4
B.3.2 Initialization of the running averages of the background noise characteristics	4
B.3.3 Generating the long-term minimum energy	5
B.3.4 Generating the difference parameters.....	5
B.3.5 Multi-boundary initial voice activity decision	6
B.3.6 Voice activity decision smoothing	7
B.3.7 Updating the running averages of the background noise characteristics.....	8
B.4 Detailed description of the DTX/CNG algorithms.....	9
B.4.1 Description of the DTX algorithm	9
B.4.2 SID evaluation and quantization	11
B.4.3 SID bit stream description.....	13
B.4.4 Non-active encoder/decoder (CNG) description	13
B.4.5 Frame erasure concealment with regards to the CNG.....	15
B.5 Bit-exact description of the silence compression scheme	15
B.5.1 Organization of the simulation software	16

A SILENCE COMPRESSION SCHEME FOR G.729 OPTIMIZED FOR TERMINALS CONFORMING TO RECOMMENDATION V.70

(Geneva, 1996)

B.1 Introduction

This annex provides a high level description of the Voice Activity Detection (VAD), Discontinuous Transmission (DTX), and Comfort Noise Generator (CNG) algorithms. These algorithms are used to reduce the transmission rate during silence periods of speech. They are designed and optimized to work in conjunction with Recommendation V.70. Recommendation V.70 mandates the use of Annex A/G.729 (G.729A) speech coding methods. However, when it is desirable, the full version of Recommendation G.729 can also be used to improve the quality of the speech. The algorithms are adapted to operate with both the full version of Recommendation G.729 and Annex A/G.729. This description is for the full version of Recommendation G.729, the only difference for Annex A is indicated in B.3.1.1. A block diagram of a silence compression speech communication system is depicted in Figure B.1.

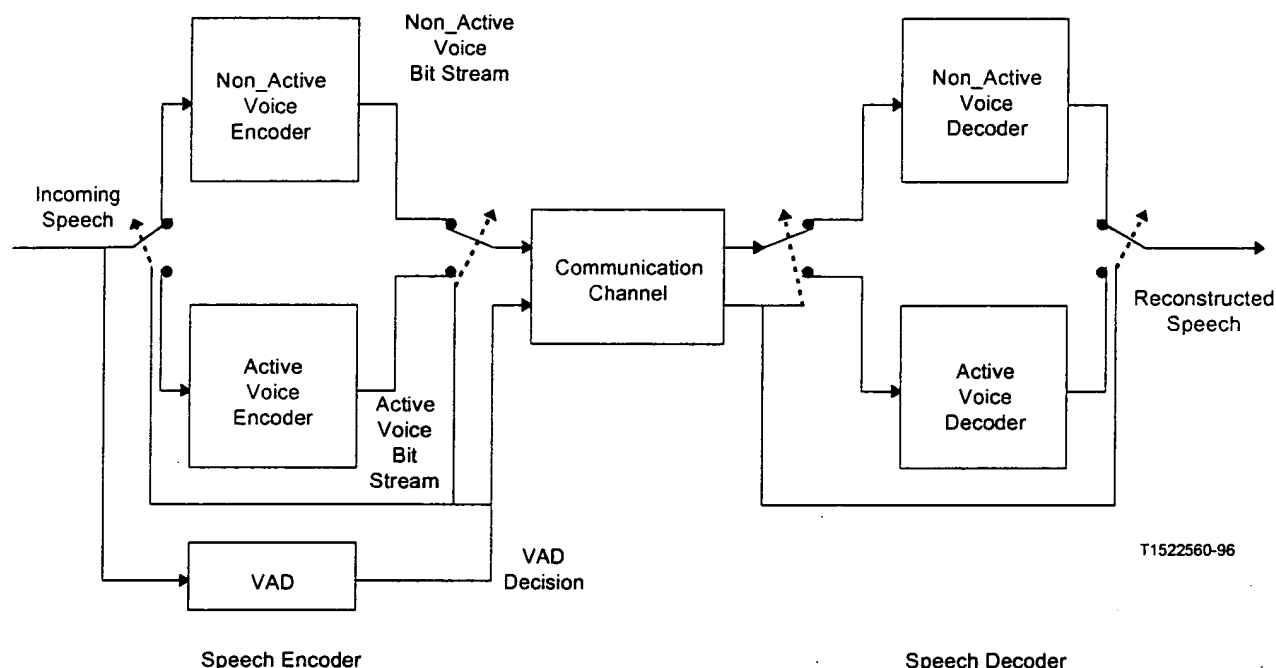


FIGURE B.1/G.729

Speech communication system with VAD

B.2 General description of the VAD/DTX/CNG algorithms

The VAD algorithm makes a voice activity decision every 10 ms in accordance with the frame size of the G.729 speech coder. A set of difference parameters is extracted and used for an initial decision. The parameters are the full band energy, the low band energy, the zero-crossing rate and a spectral measure. The long-term averages of the parameters during non-active voice segments follow the changing nature of the background noise. A set of differential parameters is obtained at each frame. These are a difference measure between each parameter and its respective long-term average.

The initial voice activity decision is obtained using a piecewise linear decision boundary between each pair of differential parameters. A final voice activity decision is obtained by smoothing the initial decision.

The output of the VAD module is either 1 or 0, indicating the presence or absence of voice activity respectively. If the VAD output is 1, the G.729 speech codec is invoked to code/decode the active voice frames. However, if the VAD output is 0, the DTX/CNG algorithms described herein are used to code/decode the non-active voice frames. Traditional speech coders and decoders use comfort noise to simulate the background noise in the non-active voice frames. If the background noise is not stationary, a mere comfort noise insertion does not provide the naturalness of the original background noise. Therefore it is desirable to intermittently send some information about the background noise in order to obtain a better quality when non-active voice frames are detected. The coding efficiency of the non-active voice frames can be achieved by coding the energy of the frame and its spectrum with as few as fifteen bits. These bits are not automatically transmitted whenever there is a non-active voice detection. Rather, the bits are transmitted only when an appreciable change has been detected with respect to the last transmitted non-active voice frame.

At the decoder side, the received bit stream is decoded. If the VAD output is 1, the G.729 decoder is invoked to synthesize the reconstructed active voice frames. If the VAD output is 0, the CNG module is called to reproduce the non-active voiced frames.

B.3 Detailed description of the VAD algorithm

A flowchart of the VAD operation is given in Figure B.2. The VAD operates on frames of digitized speech. The frames are processed in time order and are consecutively numbered from the beginning of each conversation/recording.

At the first stage, four parametric features are extracted from the input signal. Extraction of the parameters is shared with the active voice encoder module and the non-active voice encoder for computational efficiency. The parameters are the full and low-band frame energies, the set of Line Spectral Frequencies (LSF) and the frame zero crossing rate.

If the frame number is less than N_i , an initialization stage of the long-term averages takes place, and the voice activity decision is forced to 1 if the frame energy from the LPC analysis is above 15 dB (see equation B.1). Otherwise, the voice activity decision is forced to 0. If the frame number is equal to N_i , an initialization stage for the characteristic energies of the background noise occurs.

At the next stage a set of difference parameters are calculated. This set is generated as a difference measure between the current frame parameters and running averages of the background noise characteristics. Four difference measures are calculated:

- a spectral distortion;
- an energy difference;
- a low-band energy difference;
- a zero-crossing difference.

The initial voice activity decision is made at the next stage, using multi-boundary decision regions in the space of the four difference measures. The active voice decision is given as the union of the decision regions and the non-active voice decision is its complementary logical decision. Energy consideration, together with neighbouring past frames decisions, are used for decision smoothing.

The running averages have to be updated only in the presence of background noise, and not in the presence of speech. An adaptive threshold is tested, and the update takes place only if the threshold criterion is met.

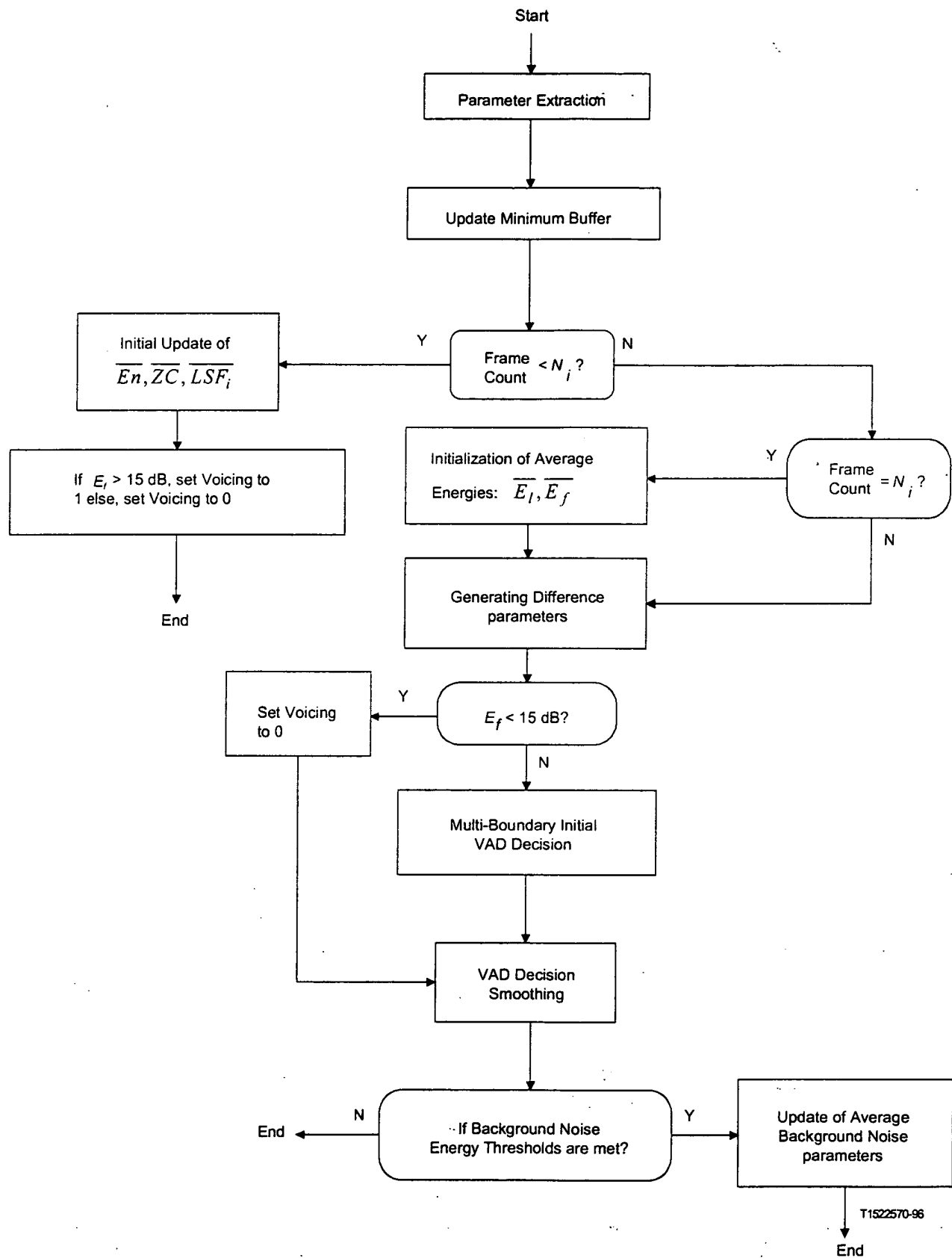


FIGURE B.2/G.729
VAD flowchart

B.3.1 Parameter extraction

For each frame a set of parameters is extracted from the speech signal. The parameters extraction module can be shared between the VAD, the active voice encoder and the non-active voice encoder. The basic set of parameters is the set of autocorrelation coefficients, which is derived similarly to Recommendation G.729 (see 3.2.1). The set of autocorrelation coefficients will be denoted by:

$$\{R(i)\}_{i=0}^q, \text{ where } q = 12$$

B.3.1.1 Line Spectral Frequencies (LSF)

A set of linear prediction coefficients is derived from the autocorrelation and a set of $\{LSF_i\}_{i=1}^p$, where $p = 10$, is derived from the set of linear prediction coefficients, as described in 3.2.3/G.729 or A.3.2.3/G.729.

B.3.1.2 Full band energy

The full band energy E_f is the logarithm of the normalized first autocorrelation coefficient $R(0)$:

$$E_f = 10 \cdot \log_{10} \left[\frac{1}{N} R(0) \right] \quad (\text{B.1})$$

where $N = 240$ is the LPC analysis window size in speech samples.

B.3.1.3 Low band energy

The low band energy E_l measured on 0 to F_l Hz band, is computed as follows:

$$E_l = 10 \cdot \log_{10} \left[\frac{1}{N} \mathbf{h}^T \mathbf{R} \mathbf{h} \right] \quad (\text{B.2})$$

where \mathbf{h} is the impulse response of an FIR filter with cutoff frequency at F_l Hz, \mathbf{R} is the Toeplitz autocorrelation matrix with the autocorrelation coefficients on each diagonal.

B.3.1.4 Zero crossing rate

Normalized zero-crossing rate ZC for each frame is calculated by:

$$ZC = \frac{1}{2M} \sum_{i=0}^{M-1} \left[\left| \text{sgn}[x(i)] - \text{sgn}[x(i-1)] \right| \right] \quad (\text{B.3})$$

where $\{x(i)\}$ is the pre-processed input signal (see 3.1/G.729) and $M = 80$.

B.3.2 Initialization of the running averages of the background noise characteristics

For the first N_i frames, the spectral parameters of the background noise, denoted by $\{\overline{LSF_i}\}_{i=1}^p$ are initialized as an average of the $\{LSF_i\}_{i=1}^p$ of the frames. The average of the background noise zero-crossings, denoted by \overline{ZC} is initialized as an average of the zero crossing rate ZC of the frames.

The running averages of the background noise energy, denoted by $\overline{E_f}$, and the background noise low-band energy, denoted by $\overline{E_l}$, are initialized as follows. First, the initialization procedure uses

$\overline{E_n}$, defined as the average of the frame energy E_f over the first N_i frames. These three averaging ($\overline{E_n}$, \overline{ZC} , and $\{\overline{LSF_i}\}_{i=1}^p$) include only the frames that have an energy E greater than 15 dB. Second, the initialization procedure continues as follows:

if $\overline{E_n} \leq T_1$ then

$$\overline{E_f} = \overline{E_n} + K_0$$

$$\overline{E_l} = \overline{E_n} + K_1$$

else if $T_1 < \overline{E_n} < T_2$ then

$$\overline{E_f} = \overline{E_n} + K_2$$

$$\overline{E_l} = \overline{E_n} + K_3$$

else

$$\overline{E_f} = \overline{E_n} + K_4$$

$$\overline{E_l} = \overline{E_n} + K_5$$

See Table B.1 for constant values.

B.3.3 Generating the long-term minimum energy

A long-term minimum energy parameter, E_{\min} , is calculated as the minimum of E_f over N_0 previous frames. Since N_0 is relatively large, E_{\min} is calculated using stored values of the minimum of E_f over short segments of the past.

B.3.4 Generating the difference parameters

Four difference measures are generated from the current frame parameters and the running averages of the background noise.

B.3.4.1 The spectral distortion ΔS

The spectral distortion measure is generated as the sum of squares of the difference between the current frame $\{LSF_i\}_{i=1}^p$ vector and the running averages of the background noise $\{\overline{LSF_i}\}_{i=1}^p$:

$$\Delta S = \sum_{i=1}^p (LSF_i - \overline{LSF_i})^2 \quad (B.4)$$

B.3.4.2 The full-band energy difference ΔE_f

The full-band energy difference measure is generated as the difference between the current frame energy, E_f , and the running average of the background noise energy, $\overline{E_f}$:

$$\Delta E_f = \overline{E_f} - E_f \quad (B.5)$$

B.3.4.3 The low-band energy difference ΔE_l

The low-band energy difference measure is generated as the difference between the current frame low-band energy, E_l , and the running average of the background noise low-band energy, $\overline{E_l}$:

$$\Delta E_l = \overline{E_l} - E_l \quad (B.6)$$

B.3.4.4 The zero-crossing difference ΔZC

The zero-crossing difference measure is generated as the difference between the current frame zero-crossing rate, ZC , and the running average of the background noise zero-crossing rate, \overline{ZC} :

$$\Delta ZC = \overline{ZC} - ZC \quad (B.7)$$

B.3.5 Multi-boundary initial voice activity decision

The initial voice activity decision is denoted by I_{VD} , and is set to 0 ("FALSE") if the vector of difference parameters lies within the non-active voice region. Otherwise, the initial voice activity decision is set to 1 ("TRUE"). The fourteen boundary decisions in the four-dimensional space are defined as follows:

- 1) if $\Delta S > a_1 \cdot \Delta ZC + b_1$ then $I_{VD} = 1$
- 2) if $\Delta S > a_2 \cdot \Delta ZC + b_2$ then $I_{VD} = 1$
- 3) if $\Delta E_f < a_3 \cdot \Delta ZC + b_3$ then $I_{VD} = 1$
- 4) if $\Delta E_f < a_4 \cdot \Delta ZC + b_4$ then $I_{VD} = 1$
- 5) if $\Delta E_f < b_5$ then $I_{VD} = 1$
- 6) if $\Delta E_f < a_6 \cdot \Delta S + b_6$ then $I_{VD} = 1$
- 7) if $\Delta S > b_7$ then $I_{VD} = 1$
- 8) if $\Delta E_l < a_8 \cdot \Delta ZC + b_8$ then $I_{VD} = 1$
- 9) if $\Delta E_l < a_9 \cdot \Delta ZC + b_9$ then $I_{VD} = 1$
- 10) if $\Delta E_l < b_{10}$ then $I_{VD} = 1$
- 11) if $\Delta E_l < a_{11} \cdot \Delta S + b_{11}$ then $I_{VD} = 1$
- 12) if $\Delta E_l > a_{12} \cdot \Delta E_f + b_{12}$ then $I_{VD} = 1$
- 13) if $\Delta E_l < a_{13} \cdot \Delta E_f + b_{13}$ then $I_{VD} = 1$
- 14) if $\Delta E_l < a_{14} \cdot \Delta E_f + b_{14}$ then $I_{VD} = 1$

If none of the fourteen conditions is "TRUE" $I_{VD} = 0$. See Table B.1 for constant values.

TABLE B.1/G.729

Table of constants

Name	Constant	Name	Constant
N_1	32	N_1	4
N_0	128	N_2	10
K_0	0	T_1	671088640
K_1	-53687091	T_2	738197504
K_2	-67108864	T_3	26843546
K_3	-93952410	T_4	40265318
K_4	-134217728	T_5	40265318
K_5	-161061274	T_6	40265318
a_1	23488	b_1	28521
a_2	-30504	b_2	19446
a_3	-32768	b_3	-32768
a_4	26214	b_4	-19661
a_5	0	b_5	-30802
a_6	28160	b_6	-19661
a_7	0	b_7	30199
a_8	16384	b_8	-22938
a_9	-19065	b_9	-31576
a_{10}	0	b_{10}	-17367
a_{11}	22400	b_{11}	-27034
a_{12}	30427	b_{12}	29959
a_{13}	-24576	b_{13}	-29491
a_{14}	23406	b_{14}	-28087

B.3.6 Voice activity decision smoothing

The initial voice activity decision is smoothed (hangover) to reflect the long-term stationarity nature of the speech signal. The smoothing is done in four stages.

A flag indicating that hangover has occurred is defined as v_flag . It is set to zero each time before the voice activity decision smoothing is performed. Denote the smoothed voice activity decision of the frame, the previous frame and frame before the previous frame by S_{VD}^0 , S_{VD}^{-1} and S_{VD}^{-2} , respectively. S_{VD}^{-1} is initialized to 1, and S_{VD}^{-2} is initialized to 1. For start $S_{VD}^0 = I_{VD}$. The first smoothing stage is:

if $(I_{VD} = 0)$ and $(S_{VD}^{-1} = 1)$ and $(E > \bar{E}_f + T3)$ then $S_{VD}^0 = 1$ and $v_flag = 1$

For the second smoothing stage define a Boolean parameter F_{VD}^{-1} and a smoothing counter C_e . F_{VD}^{-1} is initialized to 1 and C_e is initialized to 0. Denote the energy of the previous frame by E_{-1} . The second smoothing stage is:

if $(F_{VD}^{-1} = 1)$ and $(I_{VD} = 0)$ and $(S_{VD}^{-1} = 1)$ and $(S_{VD}^{-2} = 1)$ and $(|E_f - E_{-1}| \leq T_4) \{$

$$S_{VD}^0 = 1$$

$$v_flag = 1$$

$$C_e = C_e + 1$$

if $(C_e \leq N_1) \{$

$$F_{VD}^{-1} = 1$$

$\}$

else $\{$

$$F_{VD}^{-1} = 0$$

$$C_e = 0$$

$\}$

$\}$

else

$$F_{VD}^{-1} = 1$$

For the third smoothing stage define a noise continuity counter C_s , which is initialized to 0. If $S_{VD}^0 = 0$ then C_s is incremented. The third smoothing stage is:

if $(S_{VD}^0 = 1)$ and $(C_s > N_2)$ and $(E_f - E_{-1} \leq T_5) \{$

$$S_{VD}^0 = 0$$

$$C_s = 0$$

$\}$

if $(S_{VD}^0 = 1) C_s = 0$

In the fourth stage, a voice activity decision is made if the following condition is satisfied:

if $((E_f < \bar{E}_f + T_6)$ and $(frm_count > N_0)$ and $(v_flag = 0))$ then $S_{VD}^0 = 0$

B.3.7 Updating the running averages of the background noise characteristics

The running averages of the background noise characteristics are updated at the last stage of the VAD module. At this stage, the following condition is tested and the updating takes place if the following condition is met:

if $(E_f < \bar{E}_f + T_6)$ then update

The running averages of the background noise characteristics are updated using a first order Auto-regressive (AR) scheme. Different AR coefficients are used for different parameters, and different sets of coefficients are used at the beginning of the recording/conversation or when a large change of the noise characteristics is detected.

Let β_{E_f} be the AR coefficient for the update of \bar{E}_f , β_{E_l} be the AR coefficient for the update of \bar{E}_l , β_{ZC} be the AR coefficient for the update of \bar{ZC} and β_{LSF} be the AR coefficient for the update of $\{\bar{LSF}_i\}_{i=1}^p$. The total number of frames where the update condition was satisfied is counted by C_n . Different set of the coefficients β_{E_f} , β_{E_l} , β_{ZC} , and β_{LSF} is used according to the value of C_n .

The AR update is done according to:

$$\begin{aligned}\bar{E}_f &= \beta_{E_f} \cdot \bar{E}_f + (1 - \beta_{E_f}) \cdot E_f \\ \bar{E}_l &= \beta_{E_l} \cdot \bar{E}_l + (1 - \beta_{E_l}) \cdot E_l \\ \bar{ZC} &= \beta_{ZC} \cdot \bar{ZC} + (1 - \beta_{ZC}) \cdot ZC \\ \bar{LSF}_i &= \beta_{LSF} \cdot \bar{LSF}_i + (1 - \beta_{LSF}) \cdot LSF_i \quad i = 1, \dots, p\end{aligned}\tag{B.8}$$

\bar{E}_f and C_n are further updated according to:

$$\begin{aligned}&\text{if } (\text{frame count} > N_0) \text{ and } (\bar{E}_f < E_{\min}) \{ \\ &\quad \bar{E}_f = E_{\min} \\ &\quad C_n = 0 \\ &\quad \} \end{aligned}$$

B.4 Detailed description of the DTX/CNG algorithms

The DTX/CNG algorithms provide continuous and smooth information about the non-active voice periods, while keeping a low average bit rate.

B.4.1 Description of the DTX algorithm

For each non-active voice frame, the DTX module decides if a set of non-active voice update parameters ought to be sent to the speech decoder, by measuring the changes in the non-active voice signal. Absolute and adaptive thresholds on the frame energy and the spectral distortion measure are used to obtain the update decision. If an update is needed, the non-active voice encoder sends the information needed to generate a signal which is perceptually similar to the original non-active voice signal. This information is comprised of an energy level and a description of the spectral envelope. If no update is needed, the non-active voice signal is generated by the non-active decoder according to the last received energy and spectral shape information of a non-active voice frame.

However, a minimum interval of $N_{\min} = 2$ frames is required between two consecutive SID frames i.e. if a spectral or level change has occurred $n < N_{\min}$ frames after a SID frame, the SID emission is delayed.

Situated at the transmitting end, the DTX module receives from the VAD module the active/non-active voice information, and from the encoder modules the autocorrelation function of the speech signal computed for each 80 sample frame and the past excitation sample. For each frame, the DTX decision $Ftyp_t$ (Frame type for frame numbered t) is output as one of the three values, 0, 1, or 2 corresponding to untransmitted frame, active speech frame or SID frame, respectively, according to the following procedure:

B.4.1.1 Store the frame autocorrelation function

For every frame t (active or inactive), the autocorrelation coefficients of the current frame t , including the bandwidth expansion and noise correction (see the G.729 description) are retained in memory. The set of frame t autocorrelations will be denoted $r_t'(j)$, for $j = 0$ to 10.

B.4.1.2 Computation of the current frame type

If the current frame t is an active speech frame ($Vad_t = 1$), then the current frame type $Ftyp_t = 1$ and the normal speech encoder processing continues.

In the other case, a current LPC filter $A_t(z)$ calculated over $N_{cur} = 2$ previous frames including the current one t is first evaluated:

The N_{cur} autocorrelation functions are summed:

$$R^t(j) = \sum_{i=t-N_{cur}+1}^t r_i'(j), \quad j = 0 \text{ to } 10 \quad (\text{B.9})$$

and $A_t(z)$ is calculated by the Levinson-Durbin procedure (see the G.729 description) using $R^t(j)$ as input. The coefficients of this filter will be noted $a_t(j)$, $j = 0$ to 10. The Levinson-Durbin procedure also provides the residual energy E_t , that will be rescaled and used as an estimate of the frame excitation energy.

Then the current frame type $Ftyp_t$ is determined in the following way:

- If the current frame is the first inactive frame of the inactive zone, the frame is selected as SID frame. The variable \bar{E} which reflects the energy sum is taken equal to E_t , and the number of frames involved in the summation, k_E , is initialized to 1:

$$(Vad_{t-1} = 1) \Rightarrow \begin{cases} Ftyp_t = 2 \\ \bar{E} = E_t \\ k_E = 1 \end{cases} \quad (\text{B.10})$$

- For the other frames, the algorithm compares the preceding SID parameters to the current ones: if the current filter is significantly different of the preceding SID filter, or if the current excitation energy significantly differs from the preceding SID energy, the flag *flag_chang* is set to 1, else it does not change.
- The counter *count_fr* indicating how many frames are elapsed since the previous SID frame is incremented. If its value is greater than N_{min} , the emission of a SID frame is allowed. Then if *flag_chang* is equal to 1, a SID frame is sent. In all other cases, the current frame is untransmitted:

$$\left. \begin{array}{l} count_fr \geq N_{min} \\ flag_chang = 1 \end{array} \right\} \Rightarrow Ftyp_t = 2 \quad (\text{B.11})$$

Otherwise: $Ftyp_t = 0$

In case of a SID frame, the counter *count_fr* and the flag *flag_chang* are re-initialized to 0.

LPC filters and energies are compared according to the following methods:

B.4.1.3 Comparison of the LPC filters

The previous SID LPC filter will be noted $A_{sid}(z)$ and its coefficients $a_{sid}(j), j = 0$ to 10 (the evaluation of this filter is described in B.4.2.2). The current and previous SID-LPC filters are considered as significantly different if the Itakura distance between the two filters exceeds a given threshold, which is expressed by:

$$\sum_{j=0}^{10} R_a(j) \times R'(j) \geq E_t \times thr1 \quad (B.12)$$

where $R_a(j), j = 0$ to 10 is a function derived from the autocorrelation of the coefficients of the SID filter, given by:

$$\begin{cases} R_a(j) = 2 \sum_{k=0}^{10-j} a_{sid}(k) \times a_{sid}(k+j) & \text{if } j \neq 0 \\ R_a(0) = \sum_{k=0}^{10} a_{sid}(k)^2 \end{cases} \quad (B.13)$$

A value of 1.20226 is used for $thr1$.

B.4.1.4 Comparison of the energies

The sum the frame energies is calculated, k_E being first incremented up to the maximum value $Ng = 2$:

$$\bar{E} = \sum_{i=t-k_E+1}^t E_i \quad (B.14)$$

Then \bar{E} is quantized, using the 5-bits logarithmic quantizer described in B.4.2.1. The decoded log-energy E_q is compared to the previous decoded SID log-energy E_q^{sid} . If the difference exceeds the threshold $thr2 = 2$ dB, the two energies will be considered as significantly different.

B.4.2 SID evaluation and quantization

The Silence Insertion Descriptor (SID) is comprised of the quantized frame excitation energy (i.e. the current quantized excitation energy $Q(\bar{E})$ for the SID frames) and the quantized LSPs corresponding to the estimated SID-LPC filter. Four indices make up the SID frame. One index describes the energy and three indices describe the spectrum portion of the SID frame.

B.4.2.1 Energy quantization

The quantization of the energy \bar{E} is performed as follows. First, a scaling factor $\alpha_w = 0.125$ is introduced that takes into account the effect of windowing and bandwidth expansions present in the subframes autocorrelation functions $r'(j)$.

The value used at the input of the gain quantizer is:

$$E' = \alpha_w \times \frac{1}{k_E \times N_{cur} \times 80} \bar{E} \quad (B.15)$$

The energy term E' is quantized with a 5-bit non-uniform quantizer in the logarithmic domain in the range of -12 dB to 66 dB. A uniform step size of 2 dB is used between 16 dB and 66 dB. A step size

of 4 dB is used in the range of -4 dB to 16 dB. Below -4 dB, a single step size of 8 dB is used giving a quantization level of -12 dB. The quantization is straightforward and does not need the storage of a quantizer table.

Notice that since the energy comparison (B.4.1.4) is performed with decoded energies, the quantization of the energy is done for all non-active voice frames.

B.4.2.2 SID-LPC filter estimation and quantization

The SID-LPC filter estimation takes into account the local stationarity or non-stationarity of the noise at the SID frame neighbourhood.

First, a past average filter $\bar{A}_p(z)$ built from N_p frames preceding the current SID one is calculated, using the following autocorrelation sum as input of the Levinson-Durbin procedure:

$$\bar{R}_p(j) = \sum_{k=t'-N_p}^{t'} r_k'(j), \quad j = 0 \text{ to } 10 \quad (\text{B.16})$$

The number of frames involved in the summation has been fixed to $N_p = 6$.

The frame number t' varies in $[t-1, t-N_{cur}]$, depending on the rest of the Euclidian division of the current frame number t by N_{cur} .

The SID-LPC filter is then obtained with:

$$A_{sid}(z) = \begin{cases} A_t(z) & \text{if distance}(A_t(z), \bar{A}_p(z)) \geq thr3 \\ \bar{A}_p(z) & \text{otherwise} \end{cases} \quad (\text{B.17})$$

The threshold value $thr3$ is fixed to 1.12202 and the distance between the current LPC filter and the past average one is calculated in the same manner as in B.4.1.3 (see equation B.12).

Then the SID-LPC filter is transformed to the LSF domain for quantization. The LSFs are quantized by a two-stage switched predictive vector quantization ("VQ") with 5 and 4 bits each. The quantization of the LSF vector entails the determination of the best three indices. The first index is that of the predictor. The last two indices are each taken from a different vector table; as it is done in a two stage vector quantization. The overall quantization procedure follows the one given in 3.2.4/G.729 with the following modifications:

- 1) The second 4th order MA predictor used in Recommendation G.729 is modified as a linear combination of the first and second MA predictors as follows:

$$p_{i,k,2} = 0.6p_{i,k,1} + 0.4p_{i,k,2} \quad (\text{B.18})$$

where

$$i = 1, \dots, 10, \quad k = 1, \dots, 4$$

- 2) The first stage VQ quantization is similar to the one used in Recommendation G.729. However, only a portion of the first table of the quantizer is used. The relevant subset entries of the table are stored in an auxiliary lookup table with 32 address indices. Moreover, a delayed decision quantization is used by keeping few candidates as inputs to the second stage.
- 3) The candidates from the first stage in conjunction with those of the second stage are used by the second stage VQ. The second stage VQ quantization is different from the one used in Recommendation G.729. A full VQ is used as compared to the split VQ of

Recommendation G.729. Only a portion of the second stage tables is used as well. The relevant subset entries are stored in another lookup table with two 16 address entries. The combination of the predictor, a vector from the first stage and a vector from the second stage, leading to the minimum distortion in the weighted mean square error sense, is chosen as the LSF descriptor.

B.4.3 SID bit stream description

The bit stream related to the transmission of an SID frame is described in Table B.2. The bit stream related to the transmission of an active frame is defined in Table 8/G.729. The bit stream ordering is reflected by the order in the table. For each parameter the Most Significant Bit (MSB) is transmitted first.

TABLE B.2/G.729

Parameter description	Bits
Switched predictor index of LSF quantizer	1
First stage vector of LSF quantizer	5
Second stage vector of LSF quantizer	4
Gain (Energy)	5

B.4.4 Non-active encoder/decoder (CNG) description

At the decoder part, the comfort noise is generated by introducing a pseudo-white excitation signal of controlled level into interpolated LPC filters, in the same manner than the decoder produces active speech by filtering the decoded excitation. The excitation level and LPC filters are obtained from the previous SID information. The subframes interpolated LPC filters are obtained by using the SID-LSPs as current LSPs and performing the interpolation with the previous frame LSPs as done for active frames in Recommendation G.729.

The pseudo-white excitation $ex(n)$ is a mixture between an excitation of the same type as the active speech one $ex_1(n)$ and a white Gaussian excitation $ex_2(n)$.

The G.729 excitation $ex_1(n)$ is composed of an adaptive excitation with a small gain and an ACELP fixed excitation, which improves the transition between active and non-active voice frames. The addition of a Gaussian excitation $ex_2(n)$ allows the generation of a whiter signal.

Since the encoder and decoder need to keep synchronized during non-active voice periods, the excitation generation is performed on both sides, for SID frames and for untransmitted frames.

First, let us define the target excitation gain \tilde{G}_t as the square root of the average energy that must be obtained for the current frame t synthetic excitation. \tilde{G}_t is calculated using the following smoothing procedure, where \tilde{G}_{sid} is the SID gain derived for the decoded SID gain:

$$\tilde{G}_t = \begin{cases} \tilde{G}_{sid} & \text{if } Vad_{t-1} = 1 \\ \frac{7}{8}\tilde{G}_{t-1} + \frac{1}{8}\tilde{G}_{sid} & \text{otherwise} \end{cases} \quad (\text{B.19})$$

The 80 samples of the frame are divided into 2 subframes of 40 samples. For each subframe, the CNG excitation samples are synthesized using the following algorithm.

A pitch lag is randomly chosen in the interval [40,103].

Next, the fixed codebook vector of the subframe is built by random selection of the grid, the pulses signs and positions, according to the G.729 ACELP code structure.

An adaptive excitation signal of unity gain is then calculated, noted $e_a(n)$, $n = 0$ to 39. The selected subframe fixed excitation will be noted $e_f(n)$, $n = 0$ to 39.

The adaptive and fixed gains G_a and G_f are then computed in order to yield a subframe average energy equal to \tilde{G}_t^2 , which is expressed by:

$$\frac{1}{40} \sum_{n=0}^{39} (G_a \times e_a(n) + G_f \times e_f(n))^2 = \tilde{G}_t^2 \quad (\text{B.20})$$

Notice that G_f can take a negative value.

Let us define $Ea = \left(\sum_{n=0}^{39} e_a(n)^2 \right)$, $I = \left(\sum_{n=0}^{119} e_a(n) e_f(n) \right)$ and $K = 40 \times \tilde{G}_t^2$

Due to the ACELP excitation structure $\sum_{n=0}^{39} e_f(n)^2 = 4$

If we fix randomly the adaptive gain G_a , then equation B.19 becomes a second order equation on the fixed gain G_f :

$$G_f^2 + \frac{Ga \times I}{2} G_f + \frac{Ea \times Ga^2 - K}{4} = 0 \quad (\text{B.21})$$

A constraint may be imposed on G_a to be sure that this equation has a solution. Furthermore it is desirable to forbid the use of large adaptive gains. For this, the adaptive gain G_a will be randomly chosen in:

$$\left[0, \text{Max} \left\{ 0.5, \sqrt{\frac{K}{A}} \right\} \right], \text{ with } A = Ea - I^2 / 4 \quad (\text{B.22})$$

The root of equation B.20 that has the lowest absolute value is selected for G_f .

Finally the G.729 excitation is built, using:

$$ex_1(n) = G_a \times e_a(n) + G_f \times e_f[n], n = 0 \text{ to } 39 \quad (\text{B.23})$$

The method of deriving the composite excitation signal $ex(n)$ is as follows:

Let E_1 be the energy of $ex_1(n)$, E_2 be the energy of $ex_2(n)$. $ex_2(n)$ has a unit variance and a zero mean. Let E_3 be the cross-energy between $ex_1(n)$ and $ex_2(n)$.

$$\begin{aligned} E_1 &= \sum ex_1^2(n) \\ E_2 &= \sum ex_2^2(n) \\ E_3 &= \sum ex_1(n).ex_2(n) \end{aligned} \quad (\text{B.24})$$

where the summation is over the subframe size.

Let α and β be the scale proportion of $ex_1(n)$ and $ex_2(n)$ used in the mixture excitation respectively. α is set to be 0.6. β is found as the solution to the following quadratic equation:

$$\beta^2 E_2 + 2\alpha\beta E_3 + (\alpha^2 - 1)E_1 = 0, \text{ with } \beta > 0 \quad (\text{B.25})$$

If no solution is found for β , it is set to 0 and α to 1.

The CNG excitation $ex(n)$ becomes:

$$ex_1(n) = \alpha ex_1(n) + \beta ex_2(n) \quad (\text{B.26})$$

B.4.5 Frame erasure concealment with regards to the CNG

When a frame erasure is detected by the decoder, the erased frame type depends on the preceding frame type:

- if the preceding frame was active, then the current frame is considered as active;
- else if the preceding frame was either a SID frame or an untransmitted frame, the current erased frame is considered as untransmitted:

$$\begin{cases} Ftyp_{t-1} = 1 & \Rightarrow Ftyp_t = 1 \\ Ftyp_{t-1} = 0 \text{ or } 2 & \Rightarrow Ftyp_t = 0 \end{cases} \quad (\text{B.27})$$

If an untransmitted frame has been erased, no error is then introduced.

If a SID frame is erased, there are two possibilities:

- If it is not the first SID frame of the current inactive period, then the previous SID parameters are kept.
- If it is the first SID frame of an inactive period, a special protection has been taken.

Notice first that this case is detected by the fact that $Ftyp_{t-1} = 1$ and $Ftyp_t = 0$.

This combination of events does not imply that the preceding frame was a good active frame: several frames up to the preceding one may have been erased. What is certain is that the last good frame was an active frame, that the present frame was not erased, and that the SID frame supposed to provide information for the current untransmitted frame is lost.

To recover the SID information, the CNG module uses parameters provided by the G.729 decoder main part:

- the LSPs of the last valid active frame are used for the SID-LPC filter;
- an energy term is calculated on the excitation signal by the decoder during the processing of all valid active voice frames. To recover the missing SID gain \tilde{G}_{sid} , the energy term of the last valid active frame is quantized with the SID gain quantizer and decoded.

Finally to avoid de-synchronization of the random generator used to compute the excitation, the pseudo-random sequence reset is performed at each active frame, both at the encoder and decoder parts.

B.5 Bit-exact description of the silence compression scheme

The silence compression scheme is simulated in 16-bit fixed-point ANSI-C code using the same set of fixed-point basic operators defined in Table 11/G.729. The ANSI-C code constitutes an integral part of this Recommendation reflecting the bit-exact, fixed-point description of the silence compression scheme. In the event of any discrepancy between the printed text of this Recommendation and the C source, the C-source code is presumed to be correct.

B.5.1 Organization of the simulation software

Same as 5.2/G.729.

The Annex B ANSI-C software modules are listed in Table B.3. Refer to the **read.me** file provided with the software for more details.

TABLE B.3/G.729

G.729 Annex B ANSI-C module names	Description
Vad.c	VAD
Dtx.c	DTX Decision
Qsidgain.c	SID Gain Quantization
QsidLSF.c	SID-LSF Quantization
Calcexc.c	CNG Excitation Calculation
Dec_sid.c	Decode SID Information
Miscel.c	Miscellaneous Calculations
G.729 Annex B ANSI-C.h file names	Description
Vad.h	Prototype and Constants
Dtx.h	Prototype and Constants
Sid.h	Prototype and Constants
Miscel.h	Prototype and Constants

ITU-T RECOMMENDATIONS SERIES

- Series A Organization of the work of the ITU-T
- Series B Means of expression
- Series C General telecommunication statistics
- Series D General tariff principles
- Series E Telephone network and ISDN
- Series F Non-telephone telecommunication services
- Series G Transmission systems and media**
- Series H Transmission of non-telephone signals
- Series I Integrated services digital network
- Series J Transmission of sound-programme and television signals
- Series K Protection against interference
- Series L Construction, installation and protection of cables and other elements of outside plant
- Series M Maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
- Series N Maintenance: international sound-programme and television transmission circuits
- Series O Specifications of measuring equipment
- Series P Telephone transmission quality
- Series Q Switching and signalling
- Series R Telegraph transmission
- Series S Telegraph services terminal equipment
- Series T Terminal equipments and protocols for telematic services
- Series U Telegraph switching
- Series V Data communication over the telephone network
- Series X Data networks and open system communication
- Series Z Programming languages

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.